

Copy number variants in genomes of local sheep breeds from Russia

Journal:	Animal Genetics
Manuscript ID	AnGen-21-07-0221.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	14-Oct-2021
Complete List of Authors:	Igoshin, Aleksandr; The Institute of Cytology and Genetics The Siberian Branch of the Russian Academy of Sciences, Laboratory of Animal Molecular Genetics Deniskova, T; L.K Ernst Federal Science Center for Animal Husbandry, Biotecknology Yurchenko, Andrey; The Institute of Cytology and Genetics The Siberian Branch of the Russian Academy of Sciences, Laboratory of Animal Molecular Genetics Yudin, Nikolay; The Institute of Cytology and Genetics The Siberian Branch of the Russian Academy of Sciences, Laboratory of Animal Molecular Genetics Yudin, Nikolay; The Institute of Cytology and Genetics The Siberian Branch of the Russian Academy of Sciences, Laboratory of Animal Molecular Genetics Dotsev, Arsen; L.K Ernst Federal Science Center for Animal Husbandry, Biotechnology Selionova, Marina; Russian State Agrarian University Moscow Timiryazev Agricultural Academy, Russian State Agrarian University Zinovieva, Natalia A; L.K. Ernst Federal Research Center for Animal Husbandry, Biotechnology and Molecular Diagnostics Larkin, Denis; Royal Veterinary College, Comparative Biomedical Sciences; Institut citologii i genetiki SO RAN,
Keywords:	CNV, Russian sheep, traits

SCHOLARONE[™] Manuscripts

1 2 3	1	Copy number variants in genomes of local sheep breeds from Russia
4	2	
5 6	3	Alexander V. Igoshin ^{1†} , Tatiana E. Deniskova ^{2†} , Andrey A. Yurchenko ¹ , Nikolay S. Yudin ^{1,3} ,
7	4	Arsen V. Dotsev ² , Marina I. Selionova ⁴ , Natalia A. Zinovieva ^{2*} , and Denis M. Larkin ^{1,5*}
9	5	
10 11	6	¹ The Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk 630090,
12 13	7	Russia
14	8	² L K Ernst Federal Science Center for Animal Husbandry Podolsk 142132 Russia
15 16	9	³ Novosibirsk State University Novosibirsk 630090 Russia
17	10	⁴ Russian State Agrarian University, Moscow Timiryazev Agricultural Academy, Moscow
19	11	127550 Pussia
20 21	11	5 Devel Vaterinery College University of London London NW1 OTU JW
22	12	[†] These outhors contributed employ
23 24	13	These authors contributed equally.
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51		
53 54 55 56 57 58		

Summary

Copy number variants (CNVs) are genomic structural variations that contribute to many adaptive and economically important traits in livestock. In this study, we detected CNVs in 354 animals from 16 Russian indigenous sheep breeds and analysed their possible functional roles. Our analysis of the entire sample set resulted in 4,527 CNVs forming 1,450 CNV regions (CNVRs). When constructing CNVRs for individual breeds, a total of 2,715 regions ranging from 88 in Groznensk to 337 in Osetin breeds were identified. To make interbreed CNVR frequency comparison possible, we also identified core CNVRs using CNVs with overlapping chromosomal locations found in different breeds. This resulted in 137 interbreed CNVRs with frequency >15% in at least one breed. Functional enrichment analysis of genes affected by CNVRs in individual breeds revealed 12 breeds with significant enrichments in olfactory perception, PRAME family proteins, and immune response. Function of genes affected by interbreed and breed-specific CNVRs revealed candidates related to domestication, adaptation to high altitudes and cold climates, reproduction, parasite resistance, milk and meat qualities, wool traits, fat storage, and fat metabolism. Our work is the first attempt to uncover and characterize the CNV makeup of Russian indigenous sheep breeds. Further experimental and functional validation of CNVRs would help in developing new and improving existing sheep breeds.

Keywords: Local sheep breeds, CNV, CNVR, PennCNV, Russian Federation, Adaptation, eren

Economically important trait Page 3 of 28

1 2

Animal Genetics

3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30 21	
31 22	
22 22	
27	
35	
36	
37	
38	
30	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	

60

37 Introduction

38 Copy number variants (CNVs) are a type of genomic structural variation involving large-scale deletions, duplications, and insertions (Mills et al. 2011). Along with other major types of 39 40 variations, single nucleotide polymorphisms (SNPs), CNVs contribute to a substantial part of 41 phenotypic diversity in animals, therefore playing an important role in adaptive microevolution 42 in wild and domesticated species. Thus, polar bear evolution was marked by rapid changes in 43 gene copy number in response to the dietary shift (Rinker et al. 2019). Zhang et al. (2016) 44 demonstrated that CNVs might have contributed to phenotypic changes during yak 45 domestication and adaptation to life at high altitudes (Zhang et al. 2016). Serres-Armero et al. 46 (2017) identified 86 loci with large differences in CNVR content between dogs and wolves, 47 enriched in genes responsible for sensory perception, immune response, metabolic processes, etc. 48 (Serres-Armero et al. 2017). On the other hand, it has been shown that the majority of CNVRs 49 ascertained in domestic pigs are also segregating in wild boars, suggesting that domestication in 50 this case did not result in a substantial change in CNVR diversity (Paudel et al. 2013). 51 One of the species in which CNVs and their potential significance are being actively studied is 52 Ovis aries. There are several CNV studies in different sheep breeds (Fontanesi et al. 2011; Liu et 53 al. 2013; Hou et al. 2015; Ma et al. 2015; Zhu et al. 2016; Jenkins et al. 2016; Yan et al. 2017; 54 Ma et al. 2017; Yang et al. 2018; Di Gerlando et al. 2019). These reports point to CNV regions (CNVRs) possibly associated with economically important traits or involved in environmental 55 56 adaptations. For example, Zhu et al. (2016) compared breeds with different types of tail 57 morphology (fat-tailed, fat-rumped, and thin-tailed) and pointed to several candidate CNVRs 58 involved in fat deposition (Zhu et al. 2016). Ma et al. (2017) investigated CNVs in the genome 59 of indigenous Chinese Tan sheep breed and proposed a candidate gene (DLX3) for the curly 60 fleece phenotype (Ma et al. 2017). Di Gerlando and colleagues (2019) revealed association 61 between CNVRs and daily milk yield, milk fat percentage, fat yield, and protein percentage in 62 the Valle del Belice sheep breed. Furthermore, the majority of CNVRs detected (whether associated or not) overlapped known QTL regions (Di Gerlando et al. 2019). 63 64 Despite multiple studies on CNVs in sheep, there are no reports on CNVs in the Russian native sheep breeds. Due to the large territory and rich history of Russia, local sheep breeds could have 65 66 a unique repertoire of CNVs which helped them adapt to various climate conditions as well as 67 shape economically important traits. The aim of this study was to identify CNVs in sixteen 68 Russian local sheep breeds and reveal functions of genes found in most common CNVRs and 69 their possible links to the economically important and adaptive traits.

71 Material and methods

73 Sample collection, DNA extraction, and SNP genotyping

A total of 354 individuals of sixteen sheep breeds from Russia (Table S1) were analysed. Of these, 314 samples from 15 breeds have been previously used in our work (Yurchenko et al. 2019). The description of the breeds studied is provided in Table 1. Tissue samples for the Lezgin, Karachaev, Karakul, Edilbai, Romanov, Russian Longhaired, Groznensk, Salsk, Volgograd, and Osetin breeds and blood samples for the Buubei, Tuva, Altai Mountain, Krasnovarsk, Baikal, and Kulundin breeds were collected from farms and breeding centres across Russia. DNA from tissue samples of ten sheep breeds were extracted using Nexttec columns (Nexttec Biotechnology GmbH, Germany) following the manufacturer's instructions. DNA from blood were isolated using cell lysis followed by phenol-chloroform extraction. DNA samples of all breeds were genotyped on the Ovine Infinium® HD SNP BeadChip (606,006 SNPs) to produce dense SNP coverage (Yurchenko et al. 2019). Sex chromosomes and unassigned SNPs were filtered out. The remaining 577,206 SNP markers were used for detection of copy number variants (CNVs) in each sample utilizing the Oar v3.1/oviAri3 reference genome assembly.

89 CNV detection

CNVs were detected using the PennCNV software (Wang et al. 2007). The software implements the hidden Markov model (HMM) with six hidden states, each representing the copy number status of the genomic region under study. The Viterbi algorithm (Viterbi 1967) reconstructs these hidden states by utilizing the genotyping signal intensity (log R Ratio: LRR) and the B allele frequency (BAF). Apart from the PennCNV, other algorithms for CNV detection (e.g. cnvPartition (Illumina developed plug-in), Birdsuite (Korn et al. 2008), QuantiSNP (Colella et al. 2007)) could be used. While using several algorithms could potentially reduce the number of false positives, this combined approach would likely lead to many true CNVs being missed as well (Pinto et al. 2011). Therefore, we chose to run the PennCNV only as its results were shown to be the most reliable both in assessing the number of copies (Marenne et al. 2011) and in the follow up confirmation of CNVs detected (Winchester et al. 2009). To reduce the number of false positives in our study we applied several filtering steps (described below). The data on the LRR and BAF for each SNP were retrieved from the raw genotypes using the GenomeStudio 2 software (Illumina, San Diego, USA). The PFB file (containing population frequencies of B alleles) was calculated using the *compile pfb.pl* script based on the genotyping results of all 354 animals. The sheep GC model file was based on the Oar v3.1/oviAri3 assembly and contained

Page 5 of 28

Animal Genetics

summed fractions of guanines (G) and cytosines (C) in the DNA sequence (corrected for gaps)

provided a better correlation with LRR in our data compared to the ±500 Kbp interval chosen in

intervals were 0.237 and 0.166, respectively). We utilized the *hhall.hmm* model file appropriate

An initial sample-level quality control has been applied using the commonly accepted (e.g. Liu

BAF drift >0.01 and waviness factor >0.05. Nine samples have been removed after using these

filters (Table S1). In addition, to avoid false positives we removed samples with an unusually

large number of CNVs following the previous publication (Silva et al., 2016). To estimate the

individual samples and found that the mean number of unique CNVs per individual increases

drastically when including samples containing >150 CNVs. It is known that unique CNVs (also

called 'singletons') are very likely to be false positives (Kader et al. 2016; Upadhyay et al. 2017;

Lee *et al.* 2020). Therefore, for detection of CNV regions (CNVRs) we used samples with ≤ 150

CNVs (Fig. S1). A total of 265 samples have passed this threshold and were used for the

Fragmented CNV calls were merged for each animal by running the *clean cnv.pl* script with

default parameters. Then, all resulting CNVs <10 Kb in length or containing <10 SNPs were

Clusters of overlapping CNVs in all animals were merged into CNVRs using the CNVRuler

of neighboring CNVs (Kim et al., 2012). A total of 4,527 CNVs from the 265 high-quality

samples were used for this step. CNVR construction was carried out separately for the "gain"

To identify CNVRs shared between breeds (interbreed CNVRs), we identified "core" regions,

counting the number of individual CNV alleles overlapping >20% of the core region length in

present in more than one breed (Fig. S2; Appendix S1). This procedure has been applied

separately for the gain and loss events. The core CNVR frequency has been calculated by

each breed. At this stage we also used 80 individuals that did not pass the CNV number

Animal Genetics

software and applying the reciprocal overlap (RO) method based on the 50% reciprocal overlap

Construction of CNVRs and interbreed CNVR frequency analysis

removed as recommended by Lin et al. (2011) and Macé et al. (2016).

maximum acceptable CNV number (threshold) per animal we calculated the number of CNVs in

et al. 2013; Zhu et al. 2016; Rao et al. 2016; Ma et al. 2017) filtering criteria: LRR SD >0.3,

 ± 100 Kbp from each SNP marker. The ± 100 Kbp proximity interval has been chosen as it

some other studies (mean absolute correlation with LRR for the ± 100 Kbp and ± 500 Kbp

for non-human data (https://github.com/WGLab/PennCNV/issues/22).

Filtering of the PennCNV output

subsequent construction of CNVRs.

and "loss" events.

2	106
3 4	107
5 6	108
7	109
8 9	110
10 11	111
12 13	112
13 14 15	113
15 16	114
17 18	115
19 20	116
21	117
22 23	118
24 25	119
26 27	120
28	121
29 30	122
31 32	123
33 34	124
35	125
36 37	126
38 39	127
40	128
41 42	129
43 44	130
45 46	131
47	132
48 49	133
50 51	134
52	135
55 54	136
55 56	137
57 58	138
59	139
60	140

threshold (150 CNV) but passed all other quality controls. This way we used high-quality samples only to detect CNVRs allowing us to minimize the number of false positives at CNVR level while the additional 80 samples were used to calculate frequencies of these CNVRs (defined using good samples only) more precisely. Indeed, comparison of CNVR frequencies before and after supplementation indicates that our approach does not affect frequencies drastically (except for the Romanov sheep which we, therefore, excluded from the interbreed analysis; Fig. S3). The whole procedure has been performed separately for the gain and loss events. Only the interbreed CNVRs with frequencies $\geq 15\%$ in at least one breed and ≥ 10 Kbp in length were used in further analyses (Table S2). The validity of our "core" method has been confirmed by quantitative real time PCR and overlap with the CNVRs defined using whole-genome resequencing data (described later; for full details see Appendix S2). Validation of CNVRs Quantitative PCR (qPCR) was used to confirm the workability of PennCNV for our data and to check the accuracy of CNV detection and CNVR construction (See Appendix S2). We also used CNVs of Tuva and Baikal breeds detected from whole-genome sequencing data (Sweet-Jones et al. 2020) to check our CNVRs. We performed a simple overlap (1 bp minimum) of coordinates of our interbreed CNVRs (Table S2) with Tuva and Baikal breed CNVs detected in the resequencing set. **Phylogenetic tree construction** Genotypes were called using GenomeStudio 2 software (Illumina, San Diego, USA). The genotyping set has been filtered using PLINK software (Purcell et al. 2007) to remove variants with a genotyping rate of less than 99% (--geno 0.01), monomorphic SNPs and individuals with > 5% of failed calls (--mind 0.05). The variants were pruned to remove tightly linked SNPs using PLINK command: --indep-pairwise 50 10 0.1. To construct a phylogenetic tree of our populations we used the Treemix software (Pickrell & Pritchard 2012) and first tested a likely

168 number of migrational events in the range from 0 to 15 using 10 bootstrap replications for each

169 value and the Romanov breed as a root. After analysis of the LogLikelihood values and

- 52 170 associated topologies for 0-15 migrational events, we chose one migrational event as the most
 - 171 likely event and built the final tree with 100 bootstrap replications. Independent replications

172 were collapsed into the consensus tree using the majority-rule method.

59 174 Differences in CNVR frequencies between breed phylogenetic groups

Animal Genetics

2	175	As our breeds belong to two well-separated phylogenetic nodes (Figure 1) we checked if
3 4	176	interbreed CNVRs could have different frequency distribution in these phylogenetic groups. Our
5 6	177	breed sets differ in sample sizes thereby violating assumptions of the <i>t</i> -test (Bland & Kerry
7 8 9	178	1998), so a direct comparison of CNVR frequencies between clusters may be inappropriate.
	179	Therefore, we performed chi-square test comparing CNV counts in chromosome pools of each
10 11	180	cluster using the "chisq.test" function in R (Ihaka & Gentleman, 1996). The contingency tables
12 13 14	181	for each CNVR were constructed by counting the number of chromosomes carrying overlapping
	182	CNVs and those without overlapping CNV for each of the two phylogenetic groups (Table S2).
16	183	We used q -value = 0.05 as a threshold for significance.
17 18	184	
19 20	185	Genes in CNVRs and functional enrichment analysis
21	186	Coordinates of genes and their exons were obtained from the Ensembl v. 95 sheep gene
22	187	annotation file (ftp://ftp.ensembl.org/pub/release-
24 25	188	95/gtf/ovis_aries/Ovis_aries.Oar_v3.1.95.gtf.gz). We looked for genes that either had a complete
26 27	189	overlap with a CNVR or alternatively at least one exon had to have a complete overlap.
28	190	Enrichment for functional categories of genes overlapping CNVRs was performed using the
29 30	191	Database for Annotation, Visualization and Integrated Discovery (DAVID) tool (Huang et al.
31 32	192	2009). We looked for: a) functional category enrichment in CNVRs found in individual breeds
33 34	193	(minimum number of individuals per CNVR was equal to two); b) CNVRs shared by multiple
35	194	breeds (CNVR frequency >15% in at least one breed but detected in at least two breeds).
36 37	195	
38 39	196	Comparison of the CNVRs in Russian sheep to the CNVRs found in other studies
40 41	197	To estimate the fraction of known CNVRs in our set we compared our data with the combined
42	198	set of ovine CNVRs reported in several previous studies (Fontanesi et al. 2011; Liu et al. 2013;
43 44	199	Hou et al. 2015; Ma et al. 2015; Jenkins et al. 2016; Zhu et al. 2016; Ma et al. 2017; Yan et al.
45 46	200	2017; Yang et al. 2018; Di Gerlando et al. 2019; Li et al. 2020; Wang et al. 2020; Cheng et al.
47	201	2020). A simple overlap with the regions found in other publications has been reported. In those
48 49	202	cases where the type of published CNVR ('gain'/'loss') was known, only the events of the same
50 51	203	type were compared. The coordinates of CNVRs were converted to the Oar_v3.1/oviAri3
52 53	204	assembly where necessary using the liftOver tool (Kuhn et al. 2013) (Table S3).
54 54	205	
55 56	206	Results
57 58	207	CNV/CNVR detection in all samples
59 60	208	To obtain an insight into the overall distribution of CNVs in the Russian sheep breeds, we
00	209	performed CNV detection using the combined set of 345 animals representing 16 breeds. We

found a total of 46,163 raw CNVs (Tables S4-S5). After merging fragmented CNV calls, the number of CNVs has been reduced to 44,200. Of these, 18,147 CNVs (12,538 losses and 5,609 gains) were >10 Kbp and contained ≥ 10 SNPs. Considering only 265 "high quality" samples, 15,217 raw calls were obtained and subsequently merged into 15,081 CNVs, of which 4,527 (1,499 losses and 3,028 gains) were left after filtering for number of SNPs and length. To avoid false calls, we limited our CNVR detection step to a set of 265 high-quality samples. A total of 1,450 CNVRs (657 losses and 793 gains) were detected in this sample set, of which 587 (206 losses and 381 gains) were found in at least two individuals from the same or different breeds (Table S6). The CNVRs ranged in size from 10 Kbp to 1.5 Mbp, with the median size of 48.0 Kbp and covered 74.5 Mbp (3% of the sheep genome). Of these, 902 (62.2%) CNVRs had a simple overlap with at least one CNVR reported by other studies confirming the robustness of our findings and suggesting that the Russian sheep genomes could also contain CNVRs not vet reported (Table S7). Breed-specific CNVs/CNVRs To establish the breed-specific profiles of CNVs in the Russian sheep breeds we used the same approach as described above for all samples but applied it to the samples from each breed individually. The number of high-quality samples per breed varied from five in the Romanov to 21 in both Russian Longhaired and Groznensk breeds. The mean number of samples was 17. Our approach resulted in 283 CNVs on average per breed ranging from 106 in Karachaev to 857 in Osetin breeds. These CNVs have been combined into a total of 2,715 CNVRs ranging from 88 in Groznensk to 337 in Osetin sheep (Table S8). Of these CNVRs 773 were present in >1 animal within a breed with the number of such CNVRs ranging from 12 in Karachaev to 162 in Osetin sheep (Fig. S4). **Inter-breed CNVRs** Application of our core CNVR detection approach to 4,525 CNVs from 265 high-quality samples resulted in 1,093 core CNVRs of which 1,065 regions were > 10 Kbp. Out of 1,065 CNVRs, 89 loss and 48 gain regions were present with \geq 15% frequency in at least one breed (Figure 2; Table S2). Losses and gains covered 5.35 Mb (0.22%) and 3.15 Mb (0.13%) of the ovine autosomes, respectively (Figure 3). The number of carriers per CNVR ranged from eight to 142, with the median of 23 animals. Out of eight CNVRs chosen for qPCR validation, five were confirmed (Fig. S5; Table S9). For the interbreed CNVRs present either in Baikal or Tuva

- breeds, almost all (except for two gain events in the Tuva breed) regions were detected in the
 - resequencing set from Sweet-Jones et al., 2020 (Table S10).

1 2	245	
3 4	246	Phylogenetic analysis
5	247	Our phylogenetic analysis demonstrated that the Russian sheep breeds (except for the Romanov,
7	248	which was used as a root) were assigned to two major nodes (Figure 1). The first node includes
8 9	249	the Kulundin, Baikal, Salsk, Krasnoyarsk, Volgograd, Groznensk, Altai Mountain, and Russian
10 11	250	Longhaired breeds, whereas the second one includes Tuva, Buubei, Edilbai, Karakul, Lezgin,
12 13	251	Osetin, and Karachaev breeds. According to Yurchenko et al. (2019) one node of Russian breeds
14	252	combines the coarse wool breeds and another the fine and semi-fine wool breeds. In our study
15 16	253	one node corresponds to the "fine/semi-fine wool" cluster, and the second to the "coarse wool"
17 18	254	cluster (Table S2). The Osetin sheep, absent from the Yurchenko et al. (2019) study was found
19 20	255	on the coarse wool node in agreement with its wool type (State Register of Breeding
21	256	Achievements Approved for Use, <u>http://gossortrf.ru/wp-</u>
22 23	257	content/uploads/2019/07/REESTR_SKOT_2019-3.pdf).
24 25	258	
26 27	259	CNVR frequency differences between phylogenetic groups of breeds
28	260	Analysis of CNVR frequency differences between two phylogenetic groups of breeds resulted in
29 30	261	30 statistically significant (q-value < 0.05) regions, of which 25 were gains and five were losses
31 32	262	(Table 2).
33 34	263	
35	264	Gene Ontology analysis for genes overlapping CNVRs
36 37	265	To identify genes that could be affected by CNVs in the Russian sheep genomes we looked for
38 39	266	overlaps between genes annotated in the sheep genome and the CNVRs. We identified two
40 41	267	groups of genes: a) genes entirely found in a CNVR ("gene-in-CNVR"), and b) genes with
42	268	incomplete overlap but with at least one full exon found within a CNVR ("partial overlap"). In
43 44	269	total, out of 1,450 CNVRs (Table S11), 1,140 regions (79%) overlapped at least one gene. Out of
45 46	270	137 high-frequency interbreed CNVRs, 126 (92%) overlapped genes (Table S12).
47 48	271	Gene ontology analysis of CNVRs from individual breeds revealed twelve breeds with 199
49	272	statistically significant terms (q-value<0.05). Only unique terms from different databases were
50 51	273	kept in our gene enrichment analyses to avoid possible biases (Table S13). The functional
52 53	274	enrichment in gain-affected genes was mainly observed in the fine-wool breeds. The gain
54	275	CNVRs were significantly enriched for Olfactory receptor activity, Olfactory transduction, Cell
56	276	membrane and G-protein coupled receptor activity terms in the Altai Mountain and Russian
57 58	277	Longhaired breeds, while the PRAME family protein term was found enriched in the Baikal,
59 60	278	Kulundin, and Krasnoyarsk breeds. The Immunoglobulin domain and Negative regulation of
~~	279	apoptotic process terms were overrepresented in the Baikal sheep only, while the Integral

component of membrane term was unique to Buubei breed. The number of terms enriched in the loss CNVRs was somewhat larger, with the most frequent terms being Olfactory receptor activity, Olfactory transduction, G-protein coupled receptor activity and Cell membrane in the Altai Mountain, Baikal, Buubei, Edilbai, Groznensk, Karakul, Krasnoyarsk, Kulundin, Lezgin, Russian Longhaired, and Osetin breeds. The terms related to MHC cluster and immune response were uniquely overrepresented in the Kulundin sheep (see Fig. S6 for gains and Fig. S7 for losses). The functional enrichment analysis using the interbreed CNVRs produced no statistically significant results, neither for losses nor gains. Genes affected by CNVRs found exclusively in Russian breeds Out of 1,450 CNVRs detected in a set of 265 samples, 548 had no overlap with CNVRs reported elsewhere (Table S7). We found many genes uniquely affected in our dataset (when considering CNVRs of the same type) (Table S14). Most of these were affected by CNVRs in a single or few individuals. Sixteen gain regions, however, had a substantial frequency and belonged to interbreed CNVRs (Table S14). Almost all of them had the highest frequencies in Osetin breed (0.17-0.64). A gain CNVR on OAR6:114,999,642-115,041,496 overlapped *MSANTD1* (in full) and HTT (partially). In goats, MSANTD1-HTT locus is likely to be related to black coat color (Wang et al. 2016). In brown colored Osetin breed (Gochiyaev et al., 2007) this gain has frequency of 0.41. KIF13B partially overlapped by a CNVR on OAR2:102,393,643-102,423,618 was previously associated with reproductive traits in pig (number of piglets born alive) (Stafuzza et al. 2019). Its highest frequencies are in Osetin (0.26) and Karachaev (0.1) sheep. The only unique region to predominate in a breed other than Osetin is OAR3:12,568,095-12,578,995. which partially overlaps RABGAP1. This CNVR has its highest frequencies in Karachaev

(0.214) and Volgograd (0.143) breeds. The *RABGAP1* gene has been linked to the ovulatory process in cattle (Li et al. 2009). The fact that most of uniquely affected genes are harbored by

Osetin sheep could in part be explained by its history. The Osetin breed originated during the post-Mongolian period taking its roots from medieval Alan breeds (Kaloev 1993; Kaloev 2004).

CNVRs with highly diverged frequencies

We identified several CNVRs with highly different frequencies between breeds. For example, a gain CNVR on OAR20:50,919,130-50,973,643 (partial overlap with EXOC2 and IRF4) had frequencies of 0.31 and 0.02 in the Osetin and Karachaev breeds, respectively and was not found in other breeds. Similarly, a unique gain CNVR on the OAR3 affecting REV1 had frequencies of 0.64 in the Osetin breed but was absent or had frequencies < 0.1 in the other breeds. The highest frequency of a loss CNVR (0.78) was observed on OAR1:239,340,761-239,411,233 in the

Page 11 of 28

Animal Genetics

1 2	315	Lezgin breed The same CNVR had frequencies < 0.25 in the other breeds. The region affected
3	316	by this CNV does not contain known genes. A loss event with a similar location has been
4 5	317	previously reported in other studies (Liu <i>et al.</i> 2013: Yan <i>et al.</i> 2017)
6 7	318	previously reported in other studies (End et al. 2015, 1 an et al. 2017).
8 9	319	Discussion
10 11 12	320	Our study is the first to identify and study CNVs in Russian sheen genomes. We used two
	321	complementary CNVR-defining approaches. The RO approach aimed at highlighting the regions
13 14	322	of breed genomes affected by structural variations. The "core" approach aimed at defining
15 16	323	CNVRs shared between breeds CNVR detection has been performed using high-quality
17	324	genotyped samples because even for the same array a technical variability may affect output of
18 19	325	CNV-calling algorithms (Eckel-Passow <i>et al.</i> 2011: Hong <i>et al.</i> 2012). Confirmation by two
20 21	326	independent (aPCR and resequencing) approaches has proven the PennCNV's workability and
22	327	validity of our results
23 24	328	It has been shown that the CNVR makeup of livestock breeds depends on multiple factors such
25 26	329	as past changes of effective population size gene flow breeding strategies and selection
27 28	330	(Upadhyay <i>et al.</i> 2017: Strillacci <i>et al.</i> 2021) The fact that more than a half of CNVRs detected
29	331	in this study overlap CNVRs reported by others further confirms that our CNVR detection
30 31	332	approaches worked and points to a shared genetic background of Russian and worldwide sheep
32 33	333	breeds. On the other hand, some novel CNVRs could be related to environmental adaptations
34 35	334	and to distinct selection patterns. In this work we used gene functional evidence from literature
36	335	to evaluate the possibility of CNVRs to contribute to economically important traits. It should
37 38	336	also be noted that most CNVRs are detected at moderate frequencies in our breed sets
39 40	337	influencing their ability to affect breed-specific phenotypes.
41 42	338	Breed-specific analysis of gene functional terms associated with CNVRs has revealed that the
43	339	olfactory function ("olfactory transduction" and "olfactory receptor activity") was most
44 45	340	commonly affected. The enrichment of olfactory related functions in genes affected by copy
46 47	341	number variation has been previously reported in sheep (Liu <i>et al.</i> 2013), pigs (Paudel <i>et al.</i>
48 40	342	2015), cattle (Sasaki <i>et al.</i> 2016; Upadhyay <i>et al.</i> 2017), rats (Guryev <i>et al.</i> 2008; Ye <i>et al.</i> 2010),
50	343	and human (Cooper <i>et al.</i> 2007; Waszak <i>et al.</i> 2010). This enrichment could be explained either
51 52	344	by positive selection (Nguyen et al. 2006) or weaker functional constraint on the number of
53 54	345	copies for genes with olfactory function (Perry 2008). In humans, olfaction is a minor
55	346	component of fitness, therefore human olfactory receptor (OR) genes tend to evolve neutrally
57	347	(Nozawa et al. 2007). In some other mammals, however, olfaction is more important, and their
58 59	348	OR genes and variation in their copy numbers could be under selection. Thus, the forest musk
60	349	deer (Moschus berezovskii), species relying on olfaction for reproductive and social

communication, have two unique OR gene clusters absent from other Artiodactyl genomes (Bu et al. 2019). In pigs (genus Sus), expressing very strong sense of smell, OR copy number variations played a significant role in the speciation (Paudel et al. 2015). In sheep, the main olfactory system is involved in control of the social behavior (Keller & Lévy 2012) and "olfactory transduction", "olfactory receptor activity" functional terms were shown to be under selection in various breeds (Ruiz-Larrañaga et al. 2018). Our study has identified CNVRs affecting multiple genes which could be involved in breed formation, adaptation, and economically important traits in Russian sheep (for more information see Appendix S3 and Table S15). Genes reported as candidates for domestication. We identified three genes which have earlier been reported in literature as candidates associated with domestication events. GMDS was partially involved in a loss CNVR in multiple sheep breeds. Previously *GMDS* was found under positive selection in the Chinese sheep breeds when compared to their wild ancestor (Wang et al. 2019). Mutations in GMDS are associated with optic nerve degeneration. As domesticated sheep have poorer vision compared to their wild counterparts (presumably due to smaller volume of visual brain structures (Kruska 1988)), this gene could be attributed to decreased vision ability in sheep due to domestication. It is interesting that according to Haliburton and colleagues (2016), *GMDS* is also a candidate gene in the control of brain development during embryogenesis. Another gene, EXOC2 involved in both partial gains and losses in our dataset was associated with the convergent events of selection in sheep and goats (Alberto *et al.* 2018). The *EXOC2* is known to control brain development in human as its mutations cause severe brain abnormalities (Bergen *et al.* 2020). The gains and losses within this gene affect multiple exons with the highest frequency of gains in the Osetin and losses in the Volgograd sheep. The last gene reported as a candidate involved in domestication is the *KIF25* which also showed signatures of independent selection in genomes of the Chinese and European pig breeds (Wang et al. 2017). In the Russian sheep set *KIF25* was subjected to a partial deletion in six breeds (Salsk, Volgograd, Edilbai, Karakul, Lezgin, and Karachaev) with the highest frequency of the event in the Karachaev sheep. Further studies are required to understand potential contribution of CNVRs in these three genes to the domestication syndrome in sheep. Genes-candidates for adaptation to local environments. Another role that CNVRs found may play is adaptation to environmental challenges. The previous studies linked the NOXA1 to adaptations to high altitudes in goats and sheep (Wang et al. 2016). NOXA1 activates NOX1 which is involved in HIF-1 response under hypoxia conditions (Goyal et al. 2004; Malec et al. 2010). The CNVR overlapping *NOXA1* had a higher frequency in the Volgograd sheep and lower frequencies in other breeds. TMEM179, also involved in hypoxia adaptation was found

Page 13 of 28

Animal Genetics

1 2	385	downregulated in the mouse Ω_2 -sensitive neurons of the carotid body compared to the Ω_2 -
3	386	insensitive cells of superior cervical ganglion (Gao <i>et al.</i> 2017). The <i>TMEM179</i> overlaps a gain
5	387	CNVR affecting at least three exons in the mountain Osetin breed reported with moderate
6 7 8 9 10	388	frequency. The <i>REV1</i> and <i>XRCC3</i> genes are known to play a role in UV-induced mutagenesis.
	389	Partial or complete deletions of the <i>REV1</i> suppress the UV-induced mutagenesis in both veasts
	390	and mammals (Larimer <i>et al.</i> 1989; Ikehata and Ono 2011) while the <i>XRCC3</i> -deficient Chinese
12	391	hamster ovary cells exhibit increased sensitivity to DNA-damaging agents such as the UV light
13 14	392	(Bishop <i>et al.</i> 1998; Winsey <i>et al.</i> 2000). These genes have been previously associated with
15 16	393	adaptation to high altitudes in the wild boar (<i>REV1</i> , Li <i>et al.</i> 2013) and lizard (<i>XRCC3</i> , Yang <i>et</i>
17	394	<i>al.</i> 2015). In our dataset a partial gain in the <i>REV1</i> was observed with a high frequency in the
19	395	Osetin sheep supported by much lower frequencies in other breeds. Frequency of <i>XRCC3</i>
20 21	396	duplication is highest in the Osetin and lower in the Salsk, Karachaev, and Edilbai sheep. This
22 23	397	high frequency of CNV events affecting <i>REV1</i> and <i>XRCC3</i> in the Osetin breed may be a
24 25	398	consequence of the management practices. The Osetin sheep is reared in high altitude areas
26	399	under a distant-pasture system, while other (including mountain breeds) are usually reared in a
27 28	400	stall-pasture system. Therefore, it is possible that REV1 and XRCC3 found in high-frequency
29 30	401	duplication CNVRs of the Osetin sheep are beneficial for this breed.
31 32	402	Meat quality traits. Four genes in our dataset were previously associated with muscle tissue
33	403	features. KNDC1 duplicated in Krasnoyarsk and Osetin breeds is involved in regulation of
34 35	404	protein phosphorylation and was upregulated in the fast-red muscle compared to the slow-red
36 37	405	muscle in sheep (Armstrong et al. 2018). A gain with the highest frequency in the Osetin breed
38 39	406	also affects the PEX14 and CASZ1 genes both from the locus previously associated with
40	407	marbling in Nellore cattle (Magalhães et al. 2016). The product of ADI1 contributes to the meat
41 42	408	colour in beef breeds (Zhang et al. 2018). A partial loss of this gene with a moderate frequency
43 44	409	was observed in Volgograd sheep, while in other breeds frequency of this event was low.
45 46	410	Milk-related traits. We found ten genes affected by CNVRs which could contribute to milk
47	411	production traits based on previous studies. These genes have been associated with milk-related
48 49 50 51	412	traits in sheep (FASN (García-Fernández et al. 2010), SH3TC1 (Sutera et al. 2019)), goats
	413	(CASZ1 (Zhang et al. 2017)) and cattle (ALG12, DUS1L, GPS1 (Palombo et al. 2018), COG4,
52 53	414	FUK (FCSK) (Zhou et al. 2019), MGMT (Cole et al. 2011), ZC3H3 (Jiang et al. 2014)). There
53 54 55 56	415	are no pure dairy breeds in our set, though several (i.e., Altai Mountain, Karachaev, Karakul,
	416	Lezgin, Groznensk, Volgograd, and Edilbai (Dmitriev & Ernst 1989)) are reared for milk
57 58	417	purposes along with other traits. Due to the lack of published data on milk-related phenotypes of
59 60	418	the breeds under this study further research on phenotypic associations with milk traits would be
	419	required.

Wool-related traits. The *PEX14* was the only gene potentially involved in wool type determination with its CNVR frequencies significantly different between "fine wool" and "coarse wool" breed clusters. This gene is associated with hair shape (hair curliness) in humans (Liu et al. 2018). The gene encodes a component of the peroxisomal import machinery. Consistent with its contribution to the wool traits, a partial gain of the PEX14 is present in high frequency in the Osetin and in low frequencies in the Edilbai, Karachaev, and Karakul breeds, and being almost absent from the fine wool breed cluster. Conclusion Due to the large territory of Russia, its rich history and diverse climate, local sheep breeds may have a unique repertoire of CNVs originating from both European and Asian sheep. We found multiple CNV regions which could have a role in local adaptation and contribute to phenotypes. Further experimental and functional validation of the role of these CNVRs would help understanding the role of copy number variations in the formation of the local sheep breeds of Russia. Acknowledgements The authors declare no conflict of interests. The work on CNV and CNVR detection in Russian sheep breeds was funded by the Russian Scientific Foundation (RSF) grant No. 19-76-20026. The genotyping of the Osetin breed has been carried out within the project No. 21-66-00007 supported by the Russian Scientific Foundation (RSF). Availability of data The raw intensity files and genotyping data for 354 animals of 16 Russian sheep breeds were submitted to figshare under DOI 10.6084/m9.figshare.16804486 and 10.6084/m9.figshare.16806781, respectively.

Animal Genetics

1		
2	447	References
3	448	Alberto F. J. Bover F. Orozco-terWengel P. et al. (2018) Convergent genomic signatures of
4 5	449	domestication in sheen and goats Nature Communications 9 813
6	450	Armstrong E Iriarte A Nicolini P De Los Santos J Ithurralde J Bielli A Bianchi G &
7	451	Peñagaricano F (2018) Comparison of transcriptomic landscapes of different lamb muscles
8	452	using RNA-Seq. <i>PloS ONE</i> 13 e0200732 Bergen N.J. Ahmed S.M. Collins F. <i>et al.</i> (2020)
9	453	Mutations in the exocyst component EXOC2 cause severe defects in human brain development
10	454	The Journal of Experimental Medicine 217 e20192040
11	455	Bishop D K Ear U Bhattacharvya A Calderone C Beckett M Weichselbaum R R &
13	456	Shinohara A. (1998) Xrcc3 is required for assembly of Rad51 complexes in vivo. <i>The Journal of</i>
14	457	Biological Chemistry 273 , 21482–8.
15	458	Bland J.M. & Kerry S.M. (1998) Statistics notes. Weighted comparison of means. British
16	459	Medical Journal 316 , 129.
17	460	Bu P., Jian Z., Koshy J., Shen Y., Yue B. & Fan Z. (2019) The olfactory subgenome and specific
18 19	461	odor recognition in forest musk deer. Animal Genetics 50, 358-66.
20	462	Cheng J., Cao X., Hanif O., Pi L., Hu L., Huang Y., Lan X., Lei C. & Chen, H. (2020)
21	463	Integrating Genome-Wide CNVs Into QTLs and High Confidence GWAScore Regions
22	464	Identified Positional Candidates for Sheep Economic Traits. Frontiers in Genetics 11, 569.
23	465	Cole J.B., Wiggans G.R., Ma L., Sonstegard T.S., Lawlor T.J., Jr Crooker B.A., Van Tassell
24 25	466	C.P., Yang J., Wang S., Matukumalli L.K. & Da, Y. (2011) Genome-wide association analysis of
25 26	467	thirty one production, health, reproduction and body conformation traits in contemporary U.S.
20	468	Holstein cows. BMC Genomics 12, 408.
28	469	Colella S., Yau C., Taylor J.M. et al. (2007) QuantiSNP: an Objective Bayes Hidden-Markov
29	470	Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic
30	471	<i>Acids Research</i> 35 , 2013–25.
31	472	Cooper G.M., Nickerson D.A. & Eichler, E. E. (2007) Mutational and selective effects on copy-
32 33	473	number variants in the human genome. Nature Genetics 39, S22-S29.
34	474	Di Gerlando R., Sutera A.M., Mastrangelo S., Tolone M., Portolano B., Sottile G., Bagnato A.,
35	475	Strillacci M.G. & Sardina M.T. (2019) Genome-wide association study between CNVs and milk
36	476	production traits in Valle del Belice sheep. PloS ONE 14, e0215204.
37	477	Dmitriev N.G. & Ernst L.K. (1989) Animal Genetic Resources of the USSR. FAO, Rome, 517
38	478	pp.
39 40	479	Dunin I.M. & Dankvert A.G. (2013) Breeds and types of farm animals in the Russian Federation.
41	480	Ministry of Agriculture. Moscow, Russian Federation
42	481	Eckel-Passow J.E., Atkinson E.J., Maharjan S., Kardia S.L. & de Andrade, M. (2011) Software
43	482	comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array
44	483	platform. BMC Bioinformatics 12, 220.
45 46	484	Fontanesi L., Beretti F., Martelli P. L. <i>et al.</i> (2011) A first comparative map of copy number
40	485	variations in the sheep genome. Genomics 97, 158–65.
48	486	Gao L., Bonilla-Henao V., Garcia-Flores P., Arias-Mayenco I., Ortega-Saenz P. & Lopez-
49	487	Barneo J. (2017) Gene expression analyses reveal metabolic specifications in acute O2 -sensing
50	488	chemoreceptor cells. The Journal of Physiology 595, 6091–120.
51 52	489	Garcia-Fernandez M., Gutierrez-Gil B., Garcia-Gamez E., Sanchez J.P. & Arranz, J. J. (2010)
52 53	490	1 ne identification of QTL that affect the fatty acid composition of milk on sheep chromosome
55 54	491	11. Animal Genetics 41, 324–8.
55	492	Gochiyaev H.N., Gadzniev Z.K., Seikin I.I. & Aldegonova K.D. (2007) Harakteristika shersu
56	493 404	Covel D. Weissmann N. Grimminger E. et al. (2004) Unregulation of NAD(D)U evidence 1 in
57	494 405	by a r., we issuant in., Ormininger r. et al. (2004) Opregulation of NAD(P)H oxidase 1 m
58 50	490 106	Radical Biology & Medicine 36, 1270, 88
60	490 /07	Gurvey V Saar K Adamovic T et al (2008) Distribution and functional impact of DNA conv
	/0R	number variation in the rat Nature Genetics 10 538_15
	770	$\frac{1}{1000} + \frac{1}{1000} + \frac{1}{10000} + \frac{1}{10000} + \frac{1}{10000} + \frac{1}{100000} + \frac{1}{10000000000000000000000000000000000$

- 499 Haliburton G.D., McKinsey G.L. & Pollard, K.S. (2016) Disruptions in a cluster of
 500 computationally identified enhancers near FOXC1 and GMDS may influence brain
 - computationally identified enhancers near FOXC1 and GMDS may influence brain development.
 Neurogenetics 17, 1–9.
- 5 502 Hou C.L., Meng F.H., Wang W. *et al.* (2015) Genome-wide analysis of copy number variations
 - in Chinese sheep using array comparative genomic hybridization. Small Ruminant Research 12,
 19–26.
- 504 19–26.
 505 Hong H., Xu L., Liu J. *et al.* (2012) Technical reproducibility of genotyping SNP arrays used in
 506 genome-wide association studies. *PloS ONE* 7, e44483.
- 507 Huang D., Sherman B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large
 508 gene lists using DAVID bioinformatics resources. *Nature Protocols* 4, 44–57.
- ¹³ 509 Ihaka R. & Gentleman R. (1996) R: A Language for Data Analysis and Graphics. *Journal of*
- 15 510 *Computational and Graphical Statistics* **5**, 299-314.

4

- Ikehata H. & Ono T. (2011) The mechanisms of UV mutagenesis. *Journal of Radiation Research*512 52, 115–125.
- ¹⁸ 513 Jenkins G.M., Goddard M.E., Black M.A., Brauning R., Auvray B., Dodds K.G., Kijas J.W.,
- ¹⁹ ²⁰ ²¹ 514 Cockett N. & McEwan, J. C. (2016) Copy number variants in the sheep genome detected using ¹⁹ multiple approaches. *BMC Genomics* **17**, 441.
- 516 Jiang L., Liu X., Yang J. *et al.* (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics* **15**, 1105.
- 518 Kader A., Liu X., Dong K. *et al.* (2016) Identification of copy number variations in three
- 25 519 Chinese horse breeds using 70K single nucleotide polymorphism BeadChip array. Animal
 26 520 Genetics 47, 560–9.
- 521 Kaloev B.A. (1993) Skotovodstvo narodov Severnogo Kavkaza: S drevneishikh vremen do
- 522 nachala XX veka [Cattle-raising of Peoples of North Caucasus: From Antiquity to the beginning of XX century]. Moscow: Nauka, In Russian, 231 pp.
- 524 Kaloev B.A. (2004) Osetiny: Istoriko-etnograficheskoe issledovanie [Ossetians: Historical 525 Ethnographical Studies]. Moscow: Nauka, In Russian, 471 pp.
- 526 Keller M. & Lévy, F. (2012) The main but not the accessory olfactory system is involved in the
- ³⁴ 527 processing of socially relevant chemosignals in ungulates. *Frontiers in Neuroanatomy* **6**, 39.
- Kim J.H., Hu H. J., Yim S.H., Bae J. S., Kim S. Y. & Chung, Y. J. (2012) CNVRuler: a copy
 number variation based accompanying to all *Bioinformation (Or ford Final*)
- 50
 529 number variation-based case-control association analysis tool. *Bioinformatics (Oxford, England)*,
 530
 521
 521
 521
 521
- ³⁹ 531 Korn J.M., Kuruvilla F.G., McCarroll S.A. *et al.* (2008) Integrated genotype calling and
- 40 532 association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature* 41 533 *Genetics* 40, 1253–60.
- 42
 43
 43
 44
 535
 536
 537
 537
 538
 538
 539
 539
 530
 530
 531
 532
 533
 534
 535
 535
 535
 535
 536
 537
 538
 538
 539
 539
 539
 539
 530
 530
 531
 532
 532
 533
 534
 535
 535
 535
 536
 537
 538
 538
 538
 539
 539
 539
 530
 530
 531
 532
 532
 532
 533
 534
 535
 535
 536
 537
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 538
 <
- 536 Kuhn R.M., Haussler D. & Kent W.J. (2013) The UCSC genome browser and associated tools.
 537 *Briefings in Bioinformatics* 14, 144–61.
- 47 538 Larimer F.W., Perry J.R. & Hardigree A.A. (1989) The REV1 gene of Saccharomyces
- 48 539 cerevisiae: isolation, sequence, and functional analysis. *Journal of Bacteriology* **171**, 230–7.
- ⁴⁹ 540 Lee Y.L., Bosse M., Mullaart E., Groenen M., Veerkamp R.F. & Bouwman, A. C. (2020)
- 50 541 Functional and population genetic features of copy number variations in two dairy cattle 51 542 populations. *BMC Genomics* 21, 89.
- 52 542 populations. *BMC Genomics* 21, 39.
 53 543 Li X., Yang J., Shen M., *et al.* (2020) Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications* 55 545 11, 2815.
- 56 546 Li M., Tian S., Jin L. *et al.* (2013) Genomic analyses identify distinct patterns of selection in 57 547 domesticated pigs and Tibetan wild boars *Nature Genetics* **45** 1431–8
- 547 domesticated pigs and Tibetan wild boars. *Nature Genetics* **45**, 1431–8.
- ⁵⁸ 548 Li Q., Jimenez-Krassel F., Ireland J.J. & Smith G.W. (2009) Gene expression profiling of bovine
- 549 preovulatory follicles: gonadotropin surge and prostanoid-dependent up-regulation of genes 550 pretentially linked to the sympletery presence *P* prove duction (*Combailery Freedom 1*) 127, 207, 207
 - 550 potentially linked to the ovulatory process. *Reproduction (Cambridge, England)* **137**, 297–307.

- Lin P., Hartz S.M., Wang J.C. *et al.* (2011) Copy number variation accuracy in genome-wide
- ³ 552 association studies. *Human Heredity* **71**, 141–7.
- ⁴ 553 Liu F., Chen Y., Zhu G. *et al.* (2018) Meta-analysis of genome-wide association studies
- 5 554 identifies 8 novel loci involved in shape variation of human head hair. *Human Molecular* 555 *Genetics* 27, 559–75.
- 555 Genetics 27, 559–75.
 556 Liu J., Zhang L., Xu L. *et al.* (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14, 229.
- ¹⁰ 558 Ma Q., Liu X., Pan J. *et al.* (2017) Genome-wide detection of copy number variation in Chinese
- ¹¹ 559 indigenous sheep using an ovine high-density 600 K SNP array. *Scientific Reports* 7, 912.
- ¹² 560 Ma Y., Zhang Q., Lu Z., Zhao X. & Zhang, Y. (2015) Analysis of copy number variations by
 ¹³ 561 SNP50 BeadChip array in Chinese sheep. *Genomics* 106, 295–300.
- 14 Soft Soft Soft Beauchip array in Clinics sheep. *Genomics* 100, 293–300.
 15 562 Macé A., Tuke M.A., Beckmann J.S., Lin L., Jacquemont S., Weedon M.N., Reymond A. &
- Kutalik, Z. (2016) New quality measure for SNP array based CNV detection. *Bioinformatics (Oxford, England)* 32, 3298–3305.
- ¹⁸ 565 Magalhães A.F., de Camargo G.M., Fernandes G.A. *et al.* (2016) Genome-Wide Association
 ¹⁹ 566 Study of Meat Quality Traits in Nellore Cattle. *PloS ONE* 11, e0157845.
- Solar Stady of Field Quality Fields in Norror Cather Field Physics of the Physics
 Solar Physics of the Physics of the Physics of the Physics of the Physics
 Solar Physics of the Physics
 Solar Physics of the Physics
 Solar Physics of the Physics of the Physics of the Physics of the Physics
 Solar Physics of the Physics
 Solar Physics of the Physics
 Solar Physics of the Physics
 Solar Physics of the Phy
- adgmented during intermittent hypoxia by induction of the Nri2 pathway in NOX1-express
 569 adenocarcinoma A549 cells. *Free Radical Biology & Medicine* 48, 1626–35.
- 570 Marenne G., Rodríguez-Santiago B., Closas M. G. *et al.* (2011) Assessment of copy number
 571 variation using the Illumina Infinium 1M SNP-array: a comparison of methodological
- ²⁶ 572 approaches in the Spanish Bladder Cancer/EPICURO study. *Human Mutation* **32**, 240–8.
- 573 Mills R.E., Walter K., Stewart C. *et al.* (2011) Mapping copy number variation by population574 scale genome sequencing. *Nature* 470, 59–65.
- 575 Nguyen D.Q., Webber C. & Ponting C.P. (2006) Bias of selection on human copy-number
 576 variants. *PLoS Genetics* 2, e20.
- 577 Nozawa M., Kawahara Y. & Nei, M. (2007) Genomic drift and copy number variation of
- 578 sensory receptor genes in humans. *Proceedings of the National Academy of Sciences of the* 579 United States of America 104, 20421–6.
- ³⁵
 ³⁶
 ³⁷
 ³⁸
 ³⁸
 ³⁸
 ³⁹
 ³⁵
 ³⁶
 ³⁷
 ³⁸
 ³⁸
 ³⁸
 ³⁹
 ³⁹
 ³⁹
 ³¹
 ³⁵
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁸
 ³⁷
 ³⁸
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁷
 ³⁸
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁸
 ³⁷
 ³⁸
 ³⁷
 ³⁸
 ³⁷
 ³⁸
 ³⁸
 ³⁹
 <li
- 582 polymorphism arrays. *Journal of Dairy Science* **101**, 11004–19.
- 39 583 Paudel Y., Madsen O., Megens H.J., Frantz L.A., Bosse M., Crooijmans R.P. & Groenen, M. A.
- 584 (2015) Copy number variation in the speciation of pigs: a possible prominent role for olfactory
 585 receptors. *BMC Genomics* 16, 330.
- 42 586 Paudel Y., Madsen O., Megens H.J., Frantz L.A., Bosse M., Bastiaansen J.W., Crooijmans R.P.
 43 587 Paudel Y., Madsen O., Megens H.J., Frantz L.A., Bosse M., Bastiaansen J.W., Crooijmans R.P.
- 43
 44
 44
 45
 587 & Groenen, M. A. (2013) Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics* 14, 449.
- 589 Perry G.H. (2008) The evolutionary significance of copy number variation in the human
- 47 590 genome. *Cytogenetic and Genome Research* **123**, 283–7.
- 48 591 Pickrell J.K. & Pritchard, J. K. (2012) Inference of population splits and mixtures from genome 49 592 wide allele frequency data. *PLoS Genetics* 8, e1002967.
- ⁵⁰ 593 Pinto D., Darvishi K., Shi X. *et al.* (2011) Comprehensive assessment of array-based platforms
- ⁵¹ 594 and calling algorithms for detection of copy number variants. *Nature Biotechnology* **29**, 512–20.
- 52
 53
 595 Purcell S., Neale B., Todd-Brown K., *et al.* (2007). PLINK: a tool set for whole-genome
 54
 596 association and population-based linkage analyses. *American Journal of Human Genetics* 81,
- **55 597 559–575**.
- 56 598 Rao Y.S., Li J., Zhang R. *et al.* (2016) Copy number variation identification and analysis of the
 57 599 chicken genome using a 60K SNP BeadChip. *Poultry Science* 95, 1750–6.
- ⁵⁸ 600 Rinker D.C., Specian N.K., Zhao S. & Gibbons J.G. (2019) Polar bear evolution is marked by
- ⁵⁹ 60 rapid changes in gene copy number in response to dietary shift. *Proceedings of the National*
- 602 Academy of Sciences of the United States of America 116, 13446–51.

1		
2	603	Ruiz-Larrañaga O., Langa J., Rendo F., Manzano C., Iriondo M. & Estonba A. (2018) Genomic
3	604	selection signatures in sheep from the Western Pyrenees. Genetics, Selection, Evolution : GSE
4	605	50, 9.
5	606	Sasaki S., Watanabe T., Nishimura S. & Sugimoto Y. (2016) Genome-wide identification of
0 7	607	copy number variation using high-density single-nucleotide polymorphism array in Japanese
, 8	608	Black cattle. BMC Genetics 17, 26.
9	609	Serres-Armero A., Povolotskava I.S., Ouilez J. et al. (2017) Similar genomic proportions of copy
10	610	number variation within gray wolves and modern dog breeds inferred from whole genome
11	611	sequencing BMC Genomics 18 977
12	612	Silva V H Regitano L C Geistlinger L Pértille F Giachetto P F Brassaloti R A Morosini
13	613	N.S. Zimmer R. & Coutinho L. L. (2016) Genome-Wide Detection of CNVs and Their
14	614	Association with Meat Tenderness in Nelore Cattle <i>PloS ONE</i> 11 e0157711
15	615	Strillacci M.G. Marelli S.P. Milanesi R. Zaniboni I. Punturiero C. & Cerolini S. (2021) Conv.
10	616	Number Variants in Four Italian Turkey Breeds Animals : an Open Access Journal from MDPI
17	617	11 201
19	618	Stafuzza N.B. Silva R. Fragomeni B.O. Masuda V. Huang V. Grav K. & Lourenco D. (2019)
20	610	A genome-wide single nucleotide polymorphism and convinumber variation analysis for number
21	620	of niglets horn alive <i>BMC Ganomics</i> 20 , 321
22	621	Sutera A M. Riggio V. Mastrangelo S. Di Gerlando P. Sardina M.T. Pong Wong P. Tolone
23	622	M & Portalano, P. (2010) Conomo wido association studios for milk production traits in Valla
24 25	622	del Palico shaan using reported mansuras. Animal Constitutes 50 , 211, 14
26	624	Sweet Jones I. Vurchenko A. A. Jaoshin A.V. Vudin N.S. Swein M.T. & Larkin D. M. (2020)
27	625	Pasaguanging and signatures of selection scen in two Siberian native sheep breads point to
28	626	andidate genetic variants for adaptation and accommissily important traits. <i>Animal Consting</i> 52
29	620	candidate genetic variants for adaptation and economicany important trans. Animul Genetics 52,
30	627	120-51.
31	620	Number Variation agrees European Cottle Depulations, Executions in Constitution and Functionality of Copy
32 22	629	Number Variation across European Cattle Populations. <i>Frontiers in Genetics</i> 6 , 108.
33	030	viterol A.J. (1967) Error bounds for convolutional codes and an asymptotically optimum
35	631	decoding algorithm. <i>IEEE Transactions on Information Theory</i> 13 , 260–9.
36	632	Wang J., Zou H., Chen L. <i>et al.</i> (2017) Convergent and divergent genetic changes in the genome
37	633	of Chinese and European pigs. Scientific Reports 7, 8662.
38	634	Wang K., Li M., Hadley D., Liu K., Glessner J., Grant S.F., Hakonarson H. & Bucan, M. (2007)
39	635	PennCNV: an integrated hidden Markov model designed for high-resolution copy number
40	636	variation detection in whole-genome SNP genotyping data. Genome Research 17, 1665–74.
41 42	637	Wang W., Zhang X., Zhou X. et al. (2019) Deep Genome Resequencing Reveals Artificial and
42 43	638	Natural Selection for Visual Deterioration, Plateau Adaptability and High Prolificacy in Chinese
44	639	Domestic Sheep. Frontiers in Genetics 10, 300.
45	640	Wang X., Liu J., Zhou G. <i>et al.</i> (2016) Whole-genome sequencing of eight goat populations for
46	641	the detection of selection signatures underlying production and adaptive traits. <i>Scientific Reports</i>
47	642	6 , 38932.
48	643	Wang Z., Guo J., Guo Y. et al. (2020) Genome-Wide Detection of CNVs and Association With
49 50	644	Body Weight in Sheep Based on 600K SNP Arrays. Frontiers in Genetics 11, 558.
50	645	Waszak S. M., Hasin Y., Zichner T. et al. (2010) Systematic inference of copy-number
52	646	genotypes from personal genome sequencing data reveals extensive olfactory receptor gene
53	647	content diversity. <i>PLoS Computational Biology</i> 6 , e1000988.
54	648	Winchester L., Yau C. & Ragoussis, J. (2009) Comparing CNV detection methods for SNP
55	649	arrays. Briefings in Functional Genomics & Proteomics 8, 353–66.
56	650	Winsey S.L., Haldar N.A., Marsh H.P., Bunce M., Marshall S.E., Harris A.L., Wojnarowska F.
5/	651	& Welsh K.I. (2000) A variant within the DNA repair gene XRCC3 is associated with the
50 59	652	development of melanoma skin cancer. Cancer Research 60, 5612-6.
60		

- 2 653 Yan J., Blair H.T., Liu M. *et al.* (2017) Genome-wide detection of autosomal copy number 3 654 variants in several sheep breeds using Illumina OvineSNP50 BeadChips. *Small Ruminant*
 - variants in several sheep breeds using Illumina OvineSNP50 BeadChips. *Small Ruminant Research* 155, 24–32.
- 656 Yang L., Xu L., Zhou Y., Liu M., Wang L., Kijas J.W., Zhang H., Li L. & Liu, G. E. (2018)
- 658 Yang Y., Wang L., Han J. *et al.* (2015) Comparative transcriptomic analysis revealed adaptation
- 9 659 mechanism of Phrynocephalus erythrurus, the highest altitude Lizard living in the Qinghai-Tibet
 10 660 Plateau. *BMC Evolutionary Biology* 15, 101.
- 11 661 Ye Z. Q., Niu S., Yu Y. *et al.* (2010) Analyses of copy number variation of GK rat reveal new 12 662 putative type 2 diabetes susceptibility loci. *PloS ONE* **5**, e14077.
- ¹³ 663 Yurchenko A.A., Deniskova T.E., Yudin N.S. *et al.* (2019) High-density genotyping reveals
 ¹⁵ 664 signatures of selection related to acclimation and economically important traits in 15 local sheep
- 16 665 breeds from Russia. *BMC Genomics* **20**, 294.
- 17 666 Zhang X., Wang K., Wang L. *et al.* (2016) Genome-wide patterns of copy number variation in
 18 667 the Chinese yak genome. *BMC Genomics* 17, 379.
- ¹⁹ 668 Zhang X., Zhang S., Ma L. *et al.* (2017) Reduced representation bisulfite sequencing (RRBS) of
- a dairy goat mammary glands reveals DNA methylation profiles of integrated genome-wide and critical milk-related genes. *Oncotarget* 8, 115326–44.
- ²²₂₃ 671 Zhang Y.M., Zhang X.Z., Wang T.T. *et al.* (2018) Implications of step-chilling on meat color
- 672 investigated using proteome analysis of the sarcoplasmic protein fraction of beef longissimus
 673 lumborum muscle. *Journal of Integrative Agriculture* 17, 2118-25.
- ²⁶ 674 Zhou C., Li C., Cai W., Liu S., Yin H., Shi S., Zhang Q. & Zhang, S. (2019) Genome-Wide
- Association Study for Milk Protein Composition Traits in a Chinese Holstein Population Using a
 676 Single-Step Approach. *Frontiers in Genetics* 10, 72.
- ²⁹ 677 Zhu C., Fan H., Yuan, Z. *et al.* (2016) Genome-wide detection of CNVs in Chinese indigenous
- 678 sheep with different types of tails using ovine high-density 600K SNP arrays. *Scientific Reports* 679 6, 27822.

2	681	Tables							
3	682	Table 1. Information	on breeds u	sed in this	study (Dmitriev a	& Ernst 1	989; Dun	in & Danky	vert
4	683	2013).			2		,		
5	Breed	Background	Main use	Wool type	Body sizes (live weight) Distribut	tion in 1	Population	
7	Buubei	Long-term selection of native	e Meat	Coarse	51 - 53 kg (♀), 67 - 75 kg (♂)	Far Easte	ern	15 600	
8 9	Edilbai	Kazakh fat-rumped x Astrakhan	Meat, fat	Coarse	$65 - 70 \text{ kg} (\bigcirc), 110 - 12 \text{ kg} (\checkmark)$	0 Southern	regions	98 700	
10	Lezgin	Local Dagestan sheep x Tushin	Meat-milk- wool	Coarse	$45 - 50 \text{ kg} (\bigcirc), 65 - 90 \text{ kg} (\checkmark)$	Southern	regions	115 500	
12	Karachaev	Local breed	Meat-milk- wool	Coarse	$45 - 50 \text{ kg} (\bigcirc), 65 - 90 \text{ kg} (\checkmark)$	Southern	regions in zones)	299 600	
13 14	Karakul	Ancient Central Asian breed	Pelts, meat	Coarse	45 - 50 kg (♀), 65 - 90 kg (♂)	Southern (steppe a	n regions and semi-	34 600	
15 16	Osetin	Local breed	Meat-milk-	Coarse	45 - 50 kg ($\stackrel{\bigcirc}{_+}$), 65 - 90	Southern	regions	1 300	
17 18 19	Romanov	Long-term folk selection of local sheep belonging to Northern European short- tailed breeds	Meat	Coarse	$45 - 50 \text{ kg} (\bigcirc)$ $45 (\bigcirc)$	Central r	regions	53 700	
20	Tuva	Local breed	Meat	Coarse	45 - 56 kg (\bigcirc), 67 - 78	Eastern S	Siberia	201 800	
21 22	Altai Mountain	Local coarse wool fat-tailed sheep x Groznensk x Tsigai	Wool, meat	Semi-fine	$kg(\bigcirc)$ 50 - 55 kg(\bigcirc), 85 - 90 kg(\checkmark)	Western	Siberian	68 800	
23 24	Russian Longhaired	Mikhnov x Kuchugur x northern coarse-wooled shee	Meat, wool	Semi-fine	kg (♂) 60 - 65 kg (♀), 95 - 100 kg (♂)	Central r	regions	1 400	
25 26 27	Baikal	Buryat-Mongolian x Soviet Merino x Merino Précoce x Ascanian x Groznensk x Altai	Wool, meat	Fine	52 - 58 kg (♀), 100 - 11 kg (♂)	5 Eastern S regions	Siberia	131 300	
28 29	Groznensk	New Caucasian and Mazaev x Australian Merino	Wool	Fine	45 - 52 kg (♀), 85 - 100 kg (♂)	Southern	regions	295 600	
30 31 32	Krasnoyarsk	Local coarse wool sheep x Merino Précoce x American Rambouillet x Ascanian x Groznensk	Wool, meat	Fine	55 - 60 kg (♀), 100 - 11 kg (♂)	0 Eastern S regions	Siberian	3 800	
33 34	Kulundin	Altai x Groznensk x Australian Merino x Manych Merino	Wool, meat	Fine	55 - 65 kg (♀), 100 - 13 kg (♂)	0 Western regions	Siberian	1 800	
35 36	Salsk	New Caucasian and Mazaev x American Rambouillet	Wool	Fine	50 - 55 kg (♀), 95 - 110 kg (♂)	Southern	regions	4 500	
37 38 39	Volgograd	Local coarse wool sheep x Merino Précoce x Caucasian x Groznensk x New Caucasian	Meat, wool	Fine	kg (♂) 50 - 55 kg (♀), 95 - 110 kg (♂)	Southern	n regions	100 700	
40	684	* The number of reco	rded pure-b	red animal	ls under control of	f producti	vity acco	rding to the)
41 42	685 686	Yearbook of the All-F	Russian Res	earch Insti	tute of Animal Br	eeding (2	2019)		
43	687	Table 2. CNVRs and	overlanning	o genes sho	owing significant	frequency	v differen	ces hetwee	n
44 45	688	"fine wool" and "coar	se wool" cl	usters of b	reeds.	nequene.	, 411101011		
46 47 48 49	-	Chr: coordinates Cl ty	NVR Gene ov pe	erlap	Gene in CNVR "Fin clus freq (un ave	ne-wool" ster mean juency weighted rage)	"Coarse wool" cluster mean frequency (unweighted	p-value	q-value
50	-	1:239340761-239411233 L0	OSS		0.00)9	average) 0.175	5.70E-13	7.75E-11
51 52		2:248397045-248465285 G	AIN <i>SDHB; I</i> <i>Crocc</i>	MFAP2;	ATP13A2 0.00)9	0.062	0.0005193	0.0037171
53		3:728371-840735 GA	AIN EXD3; I	EDF1	NRARP; TRAF2 0.01	15	0.093	2.01E-05	0.0003037
54 55		3:2920023-2940748 GA	AIN ENSOAL	RG0000002	0.01	12	0.076	0.0001383	0.0012538
56 57		3:3309928-3364686 G.	5946; SI AIN ENSOA 5949; N	EC16A RG0000002 A CC2 ;	0		0.034	0.0018147	0.0107303
58 59		3:101824813-101850382 G	UBACI AIN REVI		0		0.113	5.99E-10	2.72E-08

Animal Genetics

3:164031471-164108598	LOSS		ENSOARG000000	0.247	0.053	7.22E-11	4.91E-09
3-223529804-223585899	GAIN	ENSOARG000001	18181; OR6C76 ZBED4: ALG12:	0.021	0.083	0 0006984	0 0047489
5.225527004 225505077	Gruit	9587	CRELD2;	0.021	0.005	0.0000704	0.0047409
			ENSOARG000000 26080;				
			ENSOARG000000				
4:112112340-112126497	GAIN	ENSOARG0000002 5262	ENSOARG000000 01402; LRRC61;	0	0.034	0.0018147	0.0107303
4-118042827 118073302	GAIN	WDRKO	RARRES2	0	0.045	0.0002628	0.0022336
4.110)42027-110)75502	UAIN	WDR00		0	0.045	0.0002028	0.0022550
5:5919565-5933187	GAIN	SIN3B		0	0.055	3.83E-05	0.0004735
5:8674173-8733419	LOSS		ENSOARG000000	0.031	0.082	0.0093046	0.042181
5:11746608-11789782	LOSS		04/48	0.403	0.174	2.37E-09	8.06E-08
5:39181048-39289328	LOSS		AC113348.1;	0.031	0.122	3.62E-05	0.0004735
			OR2AJ1; OR2L@: OR2L@				
6:114999642-115041496	GAIN	ENSOARG0000001	MSANTD1	0.003	0.069	1.29E-05	0.0002363
		3766; ENSOARG0000002					
(.115200540 115425(40	CADI	6619; HTT	ENSOADCOOOOO	0.002	0.029	0.0027794	0.0107(20
0:115598549-115455040	GAIN	FAMI95A	15067	0.003	0.038	0.003//84	0.019/038
6:116523914-116659719	GAIN	DGKQ*; CPLX1	GAK	0.006	0.093	5.83E-07	1.59E-05
7:99796479-99837532	GAIN	SPG11	RF00026	0.003	0.034	0.0070558	0.0330891
9:32337327-32352989	GAIN			0.009	0.054	0.0016908	0.0107303
11:33909668-33922382	GAIN	ТОРЗА		0	0.028	0.0066538	0.0323182
11 2450(47(245411(2	CADI			0	0.049	0.0001202	0.0012520
11:34506476-34541162	GAIN	MPRIP DEVI (***	C 1071	0	0.048	0.0001383	0.0012538
12:410684/4-4110843/	GAIN	PEX14**	CASZI	0.003	0.062	4.61E-05	0.0005225
18:66972148-67009973	GAIN	KLC1; PPP1R13B	XRCC3; ENSOARG000000	0.006	0.058	0.0003095	0.0023381
19:59145911-59233986	GAIN	GATA2; EEFSEC	00331	0.021	0.109	1.39E-05	0.0002363
20:50919130-50973643	GAIN	EXOC2; IRF4		0	0.048	0.0001383	0.0012538
22:47990537-48132672	GAIN	MGMT; EBF3		0	0.031	0.003469	0.0188714
23:41990461-42020922	GAIN	RALBP1 ; PPP4R1		0.003	0.069	1.29E-05	0.0002363
24:1312146-1373595	GAIN	MAPK8IP3; HAGH	NME3; ENSOARG000000 16416; EME2; SPSB3; NUBP2; IGFALS	0.009	0.066	0.0002858	0.0022861
24.1951409-1984366	GAIN	TEDC2; TBC1D24	NTN3	0	0.031	0.003469	0.0188714
24.1751407 1704500							

*In bold, the genes described in text are shown. **Gene is related to wool traits.





- Figure 2. Interbreed CNVRs for gains (A) and losses (B) in 15 Russian sheep breeds. On the top, phylogeny of the breeds is shown. To the left of each graph coordinates of individual CNVRs are shown. To the right, genes within CNVRs with at least one exon being affected are shown. Complete list of genes within the CNVR regions is available from Table S12.

Animal Genetics





1		
2	705	SI legends
3	706	
4	707	Figure S1. Graph showing the fraction of CNV singletons (unique CNVs in each animal) in
5	708	sheep breads in respect to the total number of CNVs detected in each animal. Numbers in blue
6 7	709	indicate animals which fall in the specific CNV ranges. Green line shows distribution for
7 8	710	duplications, red – for deletions, and brown – for all CNVs.
9	711	Figure S2 . Principle of CNVR construction using the "core" method. Boundaries of the core
10	712	CNVR are defined by the coordinates where coverage with individual CNVs from animals of all
11	712	breeds falls below 50% of the maximum coverage
12	713	Figure S3 Fold change in allele frequency for different breads after adding information for all
13	714	rigule SS. Fold change in ancie frequency for different breeds after adding information for all
14	715	annuals in the dataset.
15	/10	Figure 54. Chromosomal distribution of gain and loss CNVs for sixteen breeds using 205 nigh-
16	/1/	quality animal genotypes.
17	718	Figure S5. Five CNVRs validated by qPCR. Relative quantity of one corresponds to a diploid
18	719	state. For region chr22:50329195-5043/334, animals Baikal_18/ and Tuva_311 lacking the
20	720	CNVR from the PennCNV output but harboring it according to the resequencing data were
20	721	added to resolve the ambiguity.
22	722	Figure S6. Venn diagram for functional terms enriched in genes from gain regions.
23	723	Figure S7. Venn diagram for functional terms enriched in genes from loss regions.
24	724	Appendix S1. Procedure for defining "core" CNV regions.
25	725	Appendix S2. Validation of methods by qPCR and whole genome resequencing.
26	726	Appendix S3. Information on candidate genes.
27	727	Table S1. Samples used in the study and the corresponding PennCNV filtering criteria.
28	728	Table S2. Interbreed CNVRs with frequencies $\geq 15\%$ in at least one breed which were
29	729	constructed using the "core" method.
31	730	Table S3. Previous studies reporting copy number variations in sheep and details of coordinate
32	731	processing.
33	732	Table S4. 46,163 raw CNVs found in a set of 345 animals from 16 breeds.
34	733	Table S5. Details of CNVs in breeds
35	734	Table S6. CNVRs found in 265 high-quality samples using RO method
36	735	Table S7. Simple overlap of CNVRs found in 265 high-quality samples (Table S1) with CNVRs
37	736	from other studies
38	737	Table S8 CNVRs found in individual breeds using R0 method
39 40	738	Table S0. Statistics for aPCR confirmation and overlap with the resequencing data for eight
40	730	tasted CNVRs
42	739	Table S10 Simple overlap of interbread CNIVPs with CNIVsed data (Sweet Jones et al. 2020)
43	740	CNUPs overlepping with CNUsed/All CNUPs in group)
44	741	Table S11 Cones eventering (NVDs found in 265 high quality semples using DO method
45	742	Table S11. Genes overlapping CNVRs found in 205 nigh-quality samples using KO method.
46	743	Table S12. Genes overlapping interpreted CNVRs.
47	/44	Table S13. Statistically significant functional terms for genes found in CNVRs in individual
48	/45	breeds.
49 50	746	Table S14. CNVRs with genes uniquely affected in Russian breeds.
50	747	Table S15. Relevant functions from literature search for genes overlapping breed-specific and
52	748	interbreed CNVRs.
53	749	Table S16. Details of primers used for qPCR confirmation of CNVRs.
54		
55		
56		
5/		
50 59		





Phylogeny of the sheep breeds studied. The coarse and fine wool nodes are shown.

734x553mm (96 x 96 DPI)

2	
3	
4	
5	
6	
7	
8	
9	Unable to Convert Image
10	
11	The dimensions of this image (in pixels) are too large
12	to be converted. For this image to convert.
13	the total number of pixels (height x width) must be
14	less than 40,000,000 (40 meganizels)
15	1655 than 40,000,000 (40 megapixels).
16	
17	
18	Interbreed CNVRs for gains (A) and losses (B) in 15 Russian sheen breeds. On the top, phylogeny of the
19	breeds is shown. To the left of each graph coordinates of individual CNVRs are shown. To the right, genes
20	within CNVRs with at least one exon being affected are shown. Complete list of genes within the CNVR
21	regions is available from Table S12.
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
5/	
58	
59	Animal Constics
60	

GAIN LOSS

Animal Genetics

Number of interbreed CNVRs (>15%) per chromosome.

1972x639mm (96 x 96 DPI)

. . .



14

12

Number of CNV regions

